REMARKS

Reconsideration and allowance of the present application is respectfully requested in view of the foregoing amendments and the following additional remarks which have addressed all the issues extant in the September 7, 2004, Office Action and the January 21, 2005, Advisory action or otherwise have rendered them moot.

Claims 1-9 and 11-12, 16-18 are under consideration in this application. The claim amendments are in order to more particularly define and distinctly claim applicants' invention and/or to better recite or describe the features of the present invention as claimed. No new matter is believed to be added.

In the September 7, 2004, Office Action, the Examiner rejected claims 1-15 under 35 U.S.C. § 112, first paragraph, for allegedly reciting new matter.

Also, the Examiner rejected claims 1-15 under 35 U.S.C. § 112, second paragraph, as allegedly vague and indefinite.

Further, the Examiner objected to claim 10 for apparent grammatical error and to the specification for various formal errors.

In the January 21, 2005, Advisory action, the Examiner failed to enter the amendments filed with the response to the September 7, 2004, Office Action, alleging that new matter had been added to claim 12 and that certain elements of the claims are either not enabled or else remain vague and indefinite. These and other issues will be dealt with in the following sections.

Assessment of the Completeness of the Disclosure and the Definiteness of Claim Terms of the Instant Application Should be Based on the Knowledge and Skills of a Practitioner in Bioinformatics or Computational Biology

Applicants appreciate that terms of art such as "sufficient similarity" could elicit notions of indefiniteness in the minds of a reasonable examiner, and further, that clearly and easily apprehended terms of art such as "greedy algorithm" may sound esoteric to non-practitioners in computational biology. To the extent that the disclosure of the present application speaks to persons versed in the field of computational biology, Applicants assert that those terms do not impose burdens of undue experimentation nor would they cause exertion of a practitioner's inventive skills in order to construct and use the instant invention as disclosed.

Further, Applicants assert that the phrase "fixed-length partial sequence" is merely of lexicographic convenience to the Applicants and should not impose interpretive difficulties to one who is versed in the general art area of computational nucleic acid fragment assembly. Applicants believe that the interpretive difficulties that gave rise to the Examiner's objections and/or rejections will be ameliorated if the Examiner and the Applicants share a common framework for the nucleic acid fragment assembly problem.

Nucleic acid fragment assembly (also called partial sequence assembly) is a technique that attempts to reconstruct the original nucleic acid sequence from a large number of fragments, each several hundred base-pairs long. The nucleic acid fragment assembly is needed because current technology, such as gel electrophoresis, cannot directly and accurately sequence DNA molecules longer than 1000 bases. However, most genomes are much longer. For example, a human DNA is about 3.2 billion nucleotides in length and cannot be read at once. The art deals with this limitation thusly:

First, the DNA molecule is cut at random sites to obtain fragments that can be sequenced directly. The overlapping fragments are then assembled back into the original DNA molecule. This strategy is called shotgun sequencing. Originally, the assembly of short fragments was done by hand, which is not only inefficient but also error-prone. Hence a lot of effort has been put into finding techniques to automate the shotgun sequence assembly.

The general outline of most assembly algorithms is first to create a set of candidate overlaps by examining all pairs, followed by forming an approximate layout of fragments, and finally creating a consensus sequence. More specifically, assembling nucleic acid fragments is divided into three distinct phases – the overlap phase, the layout phase and the consensus phase.

The overlap phase consists in finding the best or longest match between the suffix of one sequence and the prefix of another. All possible pairs of fragments are compared to find their similarity. Usually, the dynamic programming algorithm is used in this step to find semiglobal alignments.

The layout phase consists of finding the order of fragments based on computed similarity scores. The Examiner's attention is particularly drawn to the notion of similarity scores which is an algorithm specific – user defined score – embedded and clearly understood by the claim phrase "sufficiently similar" upon which the Examiner based some of his rejections.

The final phase – the consensus phase – derives the DNA sequence from the layout.

As the Examiner can appreciate, accurate and fast assembly is a crucial part of any nucleic acid fragment assembly methodology. The instant invention represents patentable contributions to this methodology and is applicable not only to shotgun-assembly technique, but also to the conventional cluster-assembling of DNA sequences.

With that in mind, Applicants point out that molecular biology is in the middle of a major paradigm shift - driven by computing. Although it is already an informational science in many respects, the field is rapidly becoming much more computational and analytical. However, bridging the gap between the real world of molecular biology and the precise logical nature of computer science requires an interdisciplinary perspective. Applicants will now apply that interdisciplinary perspective in dealing with the issues particularly raised by the Examiner.

Claim Rejections Under 35 U.S.C. § 112, First Paragraph

The Examiner alleged in his September 07, 2004, rejection of claims 1-15 that the claim limitation, "comparing a sequence adjacent to said first fixed-length partial sequence of said first nucleic acid base sequence with a sequence adjacent to said second fixed-length partial sequence of the second nucleic acid base sequence to be sufficiently similar via a greedy alignment algorithm," constitutes new matter. Applicants respectfully disagree and hereby traverse as follows:

Applicants respectfully contend that there is ample recitation in the specification for "greedy alignment algorithm."

> "*When it has been found that a partial sequence 106 of a certain input sequence completely matches with a sequence defined by a fixed length window 105 as a result of referring to the table, whether it is included or not in the same cluster is verified by the detailed comparison of the sequences at the overlapping portion. Then members are included in the cluster one after another, based on a* ***greedy method*** (p. 12, lines 20-26)."

> "*In this sequence comparison, a position of the exact matching whose length is between the consensus sequence and the input sequence is apparent, so that* ***a high speed algorithm described in Zhang****, Z. et al., J. Comput. Biol., 7 (1-2): 203-14, 2000 is used* (p. 16, lines 5-9)."

Further, Applicants contend that the terms "greedy algorithm" and "high speed algorithm" are identical as indicated by the title, "A Greedy Algorithm for Aligning DNA sequences" of the publication of Zhang et al., and clearly and unambiguously known in the art of nucleic acid alignment algorithms to practitioners in computational biology; and further that they do not present any undue experimental burden. In general, a greedy algorithm is a

high speed algorithm because it represents a different problem solving modality when contrasted with dynamic programming and here is why.

Algorithms to find optimal solutions to problems typically go through a sequence of steps, with a set of choices at each step. The general strategy of dynamic programming algorithms works by solving a collection of smaller sub-problems, and building a table of solved sub problems for use in solving larger problems; eventually, this process leads to an optimal solution to a problem which consists of optimal solutions to sub problems. Since sub-problems are not solved independently, this method ensures that the same computation is not repeated needlessly. Indeed, dynamic programming is usually only of use if there are many sub-problems that crop up repeatedly when solving a problem, however it can be notoriously slow and overkill if there are no repeating sub-problems.

A greedy algorithm is so named because it always makes the choice that looks best at that moment. This simple approach is taken in the hope that a locally optimal choice will lead to a globally optimal solution.

For aligning nucleic acid sequences that differ only by sequencing errors, or by equivalent errors from other sources, a greedy algorithm can be much faster than traditional dynamic programming approaches and yet produce an alignment that is guaranteed to be theoretically optimal.

Practitioners in the field of DNA alignment are familiar with such algorithms as FASTA, BLAST, BLS2SEQ, MUMer, REPuter, Mega BLAST and so on. In particular, Mega BLAST uses the greedy algorithm for nucleotide sequence alignment searches instead of traditional dynamic programming techniques. This program is optimized for aligning sequences that differ slightly as a result of sequencing or other similar "errors". When larger word size is used it is up to 10 times faster than more common sequence similarity programs. Mega BLAST is also able to efficiently handle much longer DNA sequences than the blast program of traditional BLAST algorithm. Mega BLAST is freely available via the web at the FTP site of the National Center of Bioinformatics and is very well known in the art and it is based on the seminal paper by Z. Zhang, S. Schwartz, L. Wagner, and W. Miller. *A greedy algorithm for aligning DNA sequences*. J. of Computational Biology, 7(1-2):203-214, 2000. MegaBLAST is freely available at the FTP site of the National Center of Bioinformatics at http://www.ncbi.nlm.nih.gov/blast/megablast.html.

Referring to the general overview of the nucleic acid fragment assembly problem presented above, the complained of limitation, namely, "comparing a sequence adjacent to said first fixed-length partial sequence of said first nucleic acid base sequence with a

sequence adjacent to said second fixed-length partial sequence of the second nucleic acid base sequence to be sufficiently similar via a greedy alignment algorithm," constitute the layout phase of the problem. Applicants believe that said layout phase, although most amenable to greedy algorithm, is not necessarily so limited. As such, amended independent claims 1, 3 and 5 and their associated dependent claims do not recite a limitation to greedy algorithm. A practitioner in the art understands that the alignment problem posed by comparing sequences adjacent to those sequences that are aligned with the moving window frame can be accomplished by any method known in the art, but most efficiently by a greedy algorithm. Hence, new claims 16 – 18, recite and particularly claim the greedy algorithm limitation as a preferred method for solving the alignment problem posed by the layout phase of this invention.

It is further submitted that whereas the greedy algorithm of Mega BLAST is very well known in the art, and is most easily pluggable by an ordinarily skilled practitioner in the art in order to construct and practice the present invention, it is not required that applications be burdened by obvious and well known routines such as Mega BLAST in order to meet the requirements of 35 U.S.C. § 112.

Further, the Examiner expressed enablement difficulties arising out of reference to greedy algorithm, stating that one skilled in the art would not understand how to "assess if the second nucleic acid base sequence and the first nucleic acid base sequence can or cannot be assembled." Applicants vigorously disagree with the Examiner's observation. On first and elementary principles, without even referencing the application, two nucleic acid fragments or subsequences or partial sequences or substrings are assemblable if after determining (in the moving step of this application), that they share a fixed length partial sequence in common, it is determined, as in this case, preferably by means of greedy algorithm, that the sequences adjacent to the moving window frame are similar or aligned. The degree of alignment is expressed in the term, "sufficient similarity" and Applicants have pointed out that the similarity score is a user defined, algorithm specific number based on the optimization criteria of the algorithm in question. For instance, the greedy alignment algorithm of Mega

BLAST scores alignment by *counting the number of its differences, i.e., the number of columns that do not align identical nucleotides.*

For at least the fact that the greedy algorithm is neither new matter nor non-enabled based on the perspective of a practitioner in computational biology, Applicants respectfully assert that there is no basis to further maintain the Examiner's rejections in that regard and that those rejections be withdrawn.

Claim Rejections Under 35 U.S.C. § 112, Second Paragraph

Claims 1, 3, 5 and all dependent claims therefrom stand rejected under 35 U.S.C. § 112, Second Paragraph because the limitation, "fixed-length partial sequence" is allegedly vague and indefinite. The Examiner invited Applicants to resolve this issue by particularly pointing out what defines a "fixed – length partial sequence."

Applicants assert that the phrase, "fixed-length partial sequence" is merely of lexicographic convenience and could have just as easily been termed, "fixed-length subsequence" or "fixed length subfragment" or "fixed length substrings" to more clearly convey the notion that the invention is concerned with assembling nucleic acid fragments given a soup of nucleic acid fragments or substrings or subsequences. As for guidance in choosing the length of the "fixed-length partial sequence," there is ample description in the specification from line 9 of page 13 to line 8 of page 14.

Further, claims 1, 3, and 5 and all claims dependent therefrom stand rejected under 35 U.S.C. § 112, second paragraph on the grounds that the term, "sufficiently similar" is allegedly indefinite. As has been explained above, optimization algorithms in nucleic acid alignment are based on user specified, algorithm specific similarity scores. Such alignment algorithms like Mega BLAST consider two nucleic acid sequences to be aligned if they meet a user inputed similarity threshold. Whereas "sufficiently similar," taken out of the context of computational biology, may sound indefinite, without more, it does not provoke any vagueness in the minds of a practitioner in computational molecular biology. Nevertheless, claims 1, 3, and 5 have been amended to obviate this ground for rejection. Applicants respectfully ask that this ground of rejection be withdrawn.

Further, Applicants believe that the foregoing have adequately addressed the alleged missing essential steps on the basis of "greedy alignment algorithm." Applicants maintain that Mega BLAST, the prototypical greedy alignment algorithm is well known and easily pluggable by a practitioner to perform nucleic acid fragment assembly as taught by this invention without burdening the disclosure with its details. In particular, Applicants view the layout phase of this nucleic acid fragment alignment methodology as practiceable by any nucleic acid alignment algorithm – preferably a greedy algorithm, functioning on the whole as a subroutine that is easily pluggable into the main body of this invention. Applicants therefore do not share the Examiner's assessment that the details of greedy algorithm are a missing essential step.

From the perspective of a practitioner in computational biology, and on the basis of the foregoing, it is submitted that this ground for rejection has been adequately traversed and should be withdrawn.

Finally, in the Advisory Action of January 21, 2005, the Examiner allegedly failed to see pointed support for the amendment to claim 12 "any entry in said table is removed if a number of entries sharing an identical key therein is more than a previously specified number." Applicants believe that there is pointed support for it on page 14 line 25 through page 15, line 5. Applicants respectfully request that this rejection be withdrawn.

Conclusion

Applicants believe that all the grounds for rejections and objections have been adequately traversed or rendered moot by the foregoing amendments and remarks and earnestly solicit that the instant application be sent to issue. Should there be any outstanding issues requiring discussion that would further the prosecution and allowance of the above-

referenced application, the Examiner is invited to contact the Applicant's undersigned representative at the address and telephone number indicated below.

Respectfully submitted,

Stanley P. Fisher
Registration Number 24,344

Juan Carlos A. Marquez
Registration Number 34,072

Toni-Junell Herbert
Registration Number 34,348

REED SMITH LLP
3110 Fairview Park Drive, Suite 1400
Falls Church, Virginia 22042
(703) 641-4200

March 24, 2005
SPF/JCM/TJH